Supplementary Data

Supplementary data on Methods and Materials

Whole exome sequencing

Four distantly related individuals were selected for WES (indicated in Figure1). Exome capture was performed using Agilent's SureSelect AllExon Kit. Paired-end sequencing (2x100 base pair reads) was performed on the IlluminaHiSeq2000 platform according to the manufacturers' instructions. Reads were mapped to the human reference genome sequence (assemblyGRCh37/hg19) using the Burrows-Wheeler Alignment Tool.[1] The identified genetic variants were called with the Genome analysis Tool Kit (GATK) using the following quality criteria: Phred-like consensus quality of ≥ 30, quality by depth ≥ 5, coverage of ≥ 5 and strandbias< 0.75.[2] All variants were annotated by ANNOVAR.[3] For identification and filtering of the variants with a minor allele frequency of ≤ 1%, we used dbSNP138 (http://www.ncbi.nlm.nih.gov/projects/SNP/), the 1000 Genome Project (http://www.1000genomes.org/), and the National Heart Lung and Blood Institute (NHLBI) Grand Opportunity Exome Sequencing Project (ESP; https://esp.gs.washington.edu/drupal/). For interpretation of functional and deleterious variants, the following tools were used and added to the annotation file: 1) scores from 8 prediction algorithms (PolyPhen2[4], Sorting Intolerant from Tolerant (SIFT) [5], LRT[6], MutationTaster[7], Mutation Assessor [8], FATHMM[9], Radial SVM[10]and LR[11]; 2) a scaled Combined Annotation-Dependent Depletion (CADD) score which integrates many diverse annotations into a single measure of potential functional impact for each genetic variant[12]; a scaled CADD- score of ≥10 indicates that variants are predicted to be the 10% most deleterious substitutions in the human genome, a score of greater or equal 20 indicates the 1% most deleterious and so on; and 3) prediction scores for evolutionary conservation (GERP++ RS[13], PhyloP[14]andSiPhy[15]).

Validation by SNP genotyping

Candidate Single nucleotide variants (SNVs) from WES analysis were validated in all available family members from the pedigree using iPLEX Gold assays on a MassARRAY system (AgenaBioScience, San Diego, CA, USA). The assays were based on multiplex PCR and were designed with online available Assay Design Suite software from the same company. Clustering was called using TyperAnalyzer 4.0.22.67 software (AgenaBioScience). To ensure genotyping quality, SNV calling was verified visually to ensure distinctly formed clusters. Additionally, samples with SNV call rate lower than 50% were discarded.

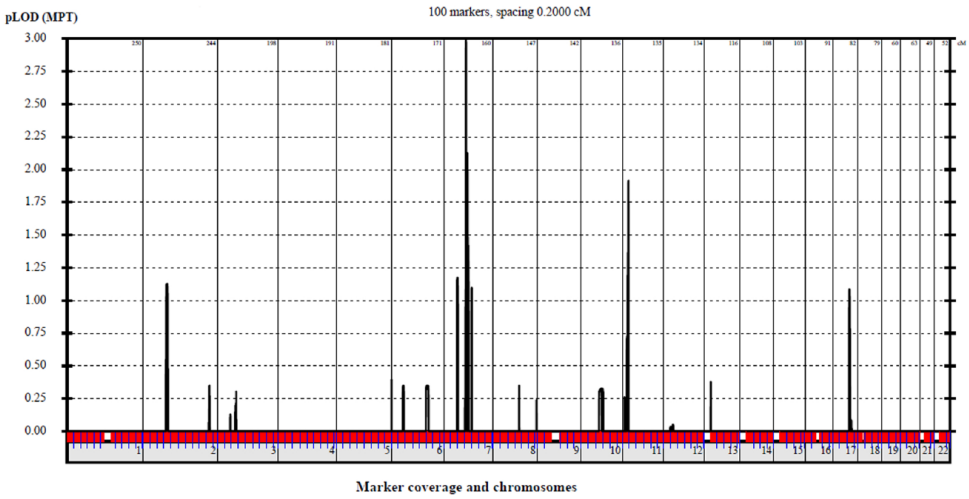
The average genotype call rate was 98.9%. Array based genotypes of independent Dutch MS cases and healthy controls from the Rotterdam study were available for a secondary analysis.

References

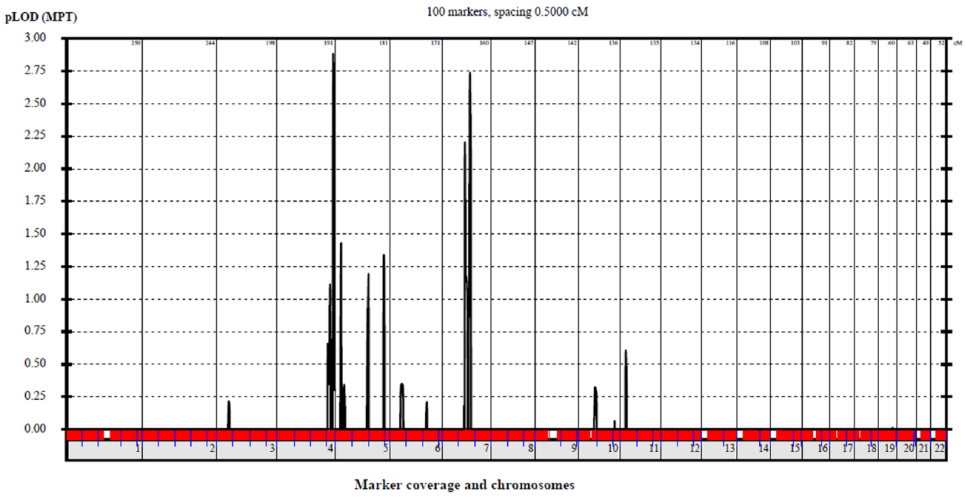
1. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754-60.
2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20: 1297-303.
3. Wang K, Li M and Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38: e164.
4. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7: 248-9.
5. Ng PC and Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001; 11: 863-74.
6. Chun S and Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009; 19: 1553-61.
7. Schwarz JM, Cooper DN, Schuelke M and Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014; 11: 361-2.
8. Reva B, Antipin Y and Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011; 39: e118.
9. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013; 34: 57-65.
10. V.V. CC. Support-vector networks. . *Mach Learn*. 1995; 20: 273-97.
11. A A. *Categorical Data Analysis*. 2002.
12. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM and Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014; 46: 310-5.
13. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A and Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010; 6: e1001025.
14. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005; 15: 901-13.
15. Garber M, Guttman M, Clamp M, Zody MC, Friedman N and Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*. 2009; 25: i54-62.

Supplementary figure 1.

A



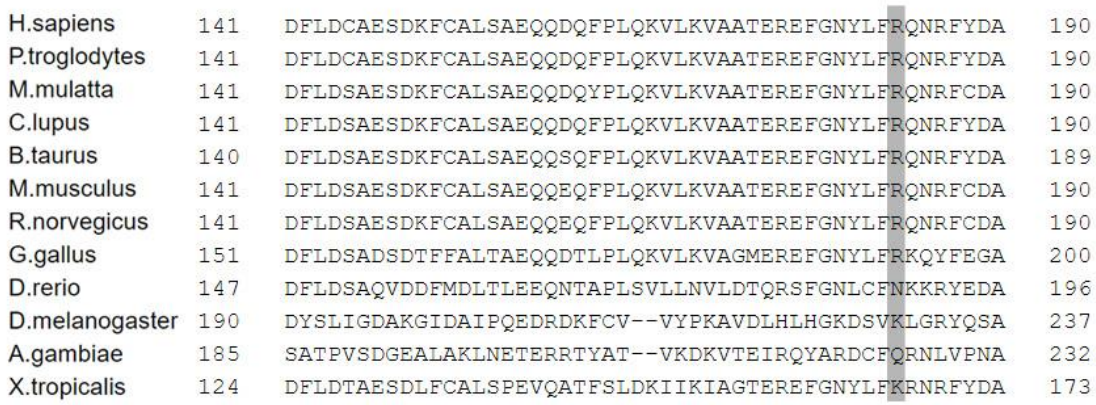
B



Supplementary figure 1 legend:

Plots of the LOD scores using Allegro. A multipoint linkage analysis was performed with a SNP spacing of 0.2 (A) and 0.5 (B) cM. Regions on the chromosome 7 with a LOD score >2.0 were used as candidate regions.

Supplementary figure 2. Sequence alignment of the FKBP6 protein across the species.



Supplementary figure 2 legend:

Multiple alignments of the FKBP6 protein across the different species. The amino acid arginine is indicated in grey and is conserved in several mammals and chicken.